DOCUMENT RESUME

ED 175 930 TM 009 669

AUTHOR Applebaum, Wayne R.
TITLE Effects of Off-Level Assignment on ITBS Grade [illegible]valent Scores.
PUB DATE Ap[illegible]79
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (63rd, San Francisco, California, April 8-12, 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Ability; Achievement Tests; Equated Scores; *Grade Equivalent Scores; *Raw Scores; Scoring; Secondary Education; Student Testing; *Test Interpretation; *Test Selection
IDENTIFIERS Dallas Independent School District TX; Iowa Tests of Basic Skills; *Out of Level Testing; *Rasch Scaled Scores; Texas (Dallas)

ABSTRACT

The Rasch model was used to compare the grade equivalent scores which students of equal ability would have obtained, had they taken different levels of the Iowa Tests of Basic Skills (ITBS). Raw scores on items common to levels 9-14 of the ITBS (vocabulary subtest) were obtained from 6,000 students in Dallas, Texas. Rasch scales were used to compare raw scores to ability (as defined by the model) on level 11, and to correlate grade equivalent scores of levels 9, 10, 12, 13 and 14 to level 11. The correlation was greater than .90. Severe discrepancies occurred, however, when a student's raw score was on the upper or lower end of the scale. It was concluded that when out-of-level testing is used and grade equivalent scores on different levels are considered on a single scale, test levels should be carefully assigned within the range of student ability. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED175930

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Effects of Off-Level Assignment on ITBS Grade Equivalent Scores

Wayne R. Applebaum

Dallas Independent School District

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Wayne R. Applebaum

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Paper presented at the annual meeting of the American Educational Research Association in San Francisco, April, 1979

TM009 669

A factor of importance in testing is that the test difficulty is within the relevant range of the student's ability. If the test is too easy or too difficult for the student, the information gained, in reference to accurately describing a student's ability, will be less than optimal. In addition, the boredom or frustration induced by a test of inappropriate difficulty may be detrimental to a student's achievement and thereby cause inaccuracies in assessment. In response to this problem, some school districts have resorted to the use of off-level testing in situations where standardized tests are used. That is, the District deviates from assigning students, in a given grade, the level of test suggested by the test publisher when there is evidence that the test is too easy or difficult for them. In such cases students are often compared across test levels by using the grade equivalent scores provided by the test publisher. A question that should be raised when off-level testing is used is, "will grade equivalent scales yield similar measures when students with the same ability take different levels of the test?" For the purpose this study ability was defined via Rasch model. The Rasch model is a probability model which states that a person with ability X has the probability of .50 of answering a question of difficulty X correctly. Thus, item difficulty and ability are measured on commensurate scales. An excellent set of articles on Latent Trait models (of which the Rasch model is one) is presented in the Journal of Educational Measurement, vol. 14, number 2.

The Rasch model allows students' abilities to be placed on a common scale if the students take a common subset of items. They need not take the same test. Levels, nine through fourteen of the Iowa Test of Basic Skills (ITBS), are organized in such a way that there are a minimum of 21 items in common on con-

secutive levels of the test. By using these common items or links, it is possible to place raw scores on all of the levels of the ITBS onto a common ability scale. Therefore, one could compare the scores of students who took different levels of the ITBS through an ability scale as well as on the grade equivalent scale. The purpose of this paper is to compare the grade equivalent scores which students with equal ability would have achieved had they taken different levels of the ITBS. This is possible because the ability scale allows raw scores and grade equivalent scores on any level to be related to each other.

Methods

The data for this study were obtained from the test records of 6,000 students who took the vocabulary subtest of the ITBS, Form 5, Levels 9-14 as part of the Dallas Independent School District's System-wide Testing program. The sample contained 1,000 students who took each level of the test. The test was administered in April of 1978.

The item scores on each level of the test were used in conjunction with the BICAL computer program (Wright and Mead, 1978) to obtain Rasch item difficulty and ability estimates. Using the procedure suggested by Wright (1977) all items and abilities were centered on a common scale. For purposes of this study, all scales equilibrated with level eleven used as a standard. Level eleven was chosen because of its convenience for further comparisons.

Results and Discussion

Table 1 shows ability levels equated to level 11 corresponding to the raw scores on each of the levels of the ITBS.

In table 2, the left-hand column gives the raw score a student received on the subtest he was given. The numbers in the remainder of the table indicate the raw score a student with the same ability would obtain had he taken

Table 1

Relation of Raw Scores to Ability on Level 11

Raw Score	Level 9	Level 10	Level 12	Level 13	Level 14
1	-4.781	-4.479	-3.555	-2.98	-2.34
2	-4.061	-3.719	-2.835	-2.24	-1.62
3	-3.621	-3.249	-2.395	-1.79	-1.18
4	-3.291	-2.899	-2.065	-1.46	-0.85
5	-3.031	-2.609	-1.805	-1.18	-0.59
6	-2.801	-2.379	-1.585	-0.95	-0.37
7	-2.591	-2.169	-1.385	-0.74	-0.17
8	-2.411	-1.979	-1.215	-0.56	0.00
9	-2.241	-1.809	-1.045	-0.39	0.06
10	-2.081	-1.649	-0.895	-0.23	0.30
11	-1.921	-1.499	-0.755	-0.08	0.44
12	-1.781	-1.359	-0.625	+0.11	0.57
13	-1.631	-1.219	-0.495	+0.19	0.69
14	-1.491	-1.089	-0.375	0.32	0.31
15	-1.351	-0.969	-0.255	0.44	.92
16	-1.211	-0.839	-0.135	0.56	1.03
17	-1.071	-0.719	-0.025	0.67	1.14
18	- .931	-0.599	+0.095	0.78	1.24
19	- .791	-0.479	+0.205	0.90	1.35
20	- .641	-0.369	0.315	1.00	1.45
21	- .481	-0.249	0.425	1.11	1.55
22	- .321	-0.129	0.525	1.22	1.64
23	- .151	-0.009	0.635	1.32	1.74
24	+ .029	+0.121	0.745	1.43	1.84
25	.239	+0.251	0.855	1.53	1.94
26	.469	.381	.965	1.63	2.04
27	.729	.511	1.075	1.74	2.14
28	1.059	.661	1.195	1.84	2.24
29	1.499	.811	1.305	1.95	2.34
30	2.219	.971	1.425	2.06	2.44
31		1.151	1.545	2.17	2.55
32		1.351	1.675	2.28	2.66
33		1.571	1.805	2.40	2.77
34		1.841	1.945	2.52	2.89
35		2.161	2.095	2.64	3.01
36		2.601	2.245	2.78	3.14
37		3.321	2.405	2.96	3.37
38			2.585	3.06	3.42
39			2.785	3.22	3.57
40			3.005	3.39	3.74
41			3.255	3.58	3.93
42			3.565	3.79	4.14
43			3.945	4.03	4.37
44			4.475	4.32	4.66
45			5.325	4.69	5.01
46				5.19	5.49
47				6.00	6.26
48					

Table 2

Relationships Between Raw Scores on Level Taken to Raw Scores on Level 11

Raw Score on Subtest Taken	Level Taken				
	9	10	12	13	14
1	0	0	1	2	4[1]
2	0	1	3	5	8
3	1	2	4	6	11
4	2	3	6	9	14
5	2	3	7	11	16
6	3	4	8	13	18
7	4	5	9	15	20
8	4	6	11	16	22
9	5	7	12	18	22
10	5	8	13	19	25
11	6	9	14	21	26
12	7	10	16	23	27
13	8	11	17	24	28
14	9	12	18	25	29
15	10	13	19	26	30
16	11	14	20	27	31
17	12	15	21	28	32
18	13	16	23	29	33
19	14	17	24	30	33
20	16	18	25	31	34
21	17	19	26	32	35
22	19	20	27	32	35
23	20	22	28	33	36
24	22	23	29	34	36
25	24	24	30	35	37
26	26	25	31	35	37
27	29	27	31	36	37
28	31	28	32	36	38
29	34	29	33	37	39
30	38	31	34	37	39
31		32	35	38	39
32		33	35	38	40
33		35	36	39	40
34		36	37	39	40
35		38	38	40	40
36		39	38	40	41
37		41	39	40	41
38			39	41	41
39			40	41	41
40			41	41	42
41			41	41	42
42			42	42	42
43			42	42+	43+
44			42+	42+	43+
45			42+	42+	43+
46				42+	43+
47				42+	43+
48				42+	43+

[1]Raw score on level eleven

level 11. Relating tables 1 and 2, table 1 indicates a student with a raw score of 24 on level 9 would have an estimated ability of .029. A student with a raw score of 9 on level 14 would have an estimated ability score of .06. In table 2, we see that both would have an expected score of 22 on level 11 based on their abilities there. Since raw scores on each of the tests can be related to ability and also to a grade equivalent score, it is possible to compare the grade equivalent score a person received on one level with the grade equivalent score he would have received had he taken another. Tables 3-7 demonstrate the relationship between grade equivalents of levels 9, 10, 12, 13, and 14 to level 11. The column labeled raw score is the raw score that the individual achieved on the level taken. The second column shows the grade equivalent score for the level taken. The third column indicates the grade equivalent score of a person with the same ability had they taken level 11. The correlation, between the grade equivalent scores on each of the pairs presented, was greater than .90. This obviously indicates a creditable relationship between the two scales, especially considering the sample based and interpolative nature of the grade equivalent scores. It also presents some support for the use of vertical equating. However, it would be a mistake to assume that because summary statistics do not indicate overall differences that local discrepancies in certain areas of the scale will not cause problems in interpretation.

The most severe discrepancies occur when a student's score is on the upper end of the scale. A student, who is able to answer about 90% of the items correctly on level 9, would obtain a grade equivalent score of 5.3. A student, with the same ability, would receive a score of 6.1 on level 11. The discrepancy widens to 1.6 years if the student is able to answer 30 items on level 9 correctly. Carrying out a similar comparison between levels 10 and 11, one finds that the gap at the 90% level is .6. This widens to 1.1 if a student can get 37 items on level ten. Therefore, although the gap

Table 3

Relationship Between Grade Equivalent Scores

Raw Score on Level 9	Level 9	Level 11
0	0.4 [1]	1.0 [2]
1	0.7	1.0
2	1.0	1.0
3	1.2	1.1
4	1.4	1.3
5	1.6	1.3
6	1.8	1.5
7	2.0	1.7
8	2.3	1.7
9	2.5	1.9
10	2.7	1.9
11	2.9	2.1
12	3.1	2.3
13	3.3	2.6
14	3.4	2.8
15	3.5	3.1
16	3.6	3.3
17	3.8	3.6
18	3.9	3.8
19	4.0	4.0
20	4.1	4.4
21	4.3	4.6
22	4.4	4.9
23	4.5	5.0
24	4.7	5.3
25	4.9	5.5
26	5.1	5.7
27	5.3	6.1
28	5.5	6.3
29	5.7	6.8
30	6.0	7.6
31	6.3	7.6

[1]Grade equivalent if the student had taken level 9.

[2]Grade equivalent if the student had taken level 11.

Table 4

Relationship Between Grade Equivalent Scores

Raw Score on Level 10	Level 10	Level 11
0	0.7	1.0
1	0.9	1.0
2	1.1	1.1
3	1.3	1.3
4	1.5	1.5
5	1.8	1.5
6	2.0	1.7
7	2.2	1.9
8	2.5	2.1
9	2.7	2.3
10	2.9	2.6
11	3.1	2.8
12	3.4	3.1
13	3.6	3.3
14	3.8	3.6
15	4.0	3.8
16	4.1	4.0
17	4.2	4.2
18	4.4	4.4
19	4.5	4.6
20	4.6	4.8
21	4.8	4.9
22	4.9	5.0
23	5.0	5.3
24	5.1	5.4
25	5.2	5.5
26	5.4	5.6
27	5.5	5.8
28	5.6	6.0
29	5.7	6.1
30	5.9	6.3
31	6.0	6.5
32	6.2	6.6
33	6.4	7.0
34	6.6	7.2
35	6.8	7.6
36	7.1	7.9
37	7.4	8.5
38	7.8	8.5

Table 5

Relationship Between Grade Equivalent Scores

Raw Score on Level 12	Level 12	Level 11
0	1.3	1.6
1	1.4	1.1
2	1.6	1.5
3	1.7	1.7
4	1.9	2.1
5	2.1	2.3
6	2.3	2.6
7	2.6	2.8
8	2.8	3.3
9	3.0	3.6
10	3.3	3.8
11	3.5	4.0
12	3.8	4.4
13	4.1	4.6
14	4.3	4.8
15	4.6	4.9
16	4.9	5.0
17	5.1	5.3
18	5.3	5.4
19	5.5	5.5
20	5.7	5.6
21	5.9	5.7
22	6.0	5.8
23	6.2	6.0
24	6.3	6.1
25	6.5	6.2
26	6.6	6.3
27	6.8	6.3
28	7.0	6.5
29	7.1	6.6
30	7.3	6.8
31	7.4	7.0
32	7.6	7.0
33	7.8	7.2
34	8.0	7.4
35	8.2	7.6
36	8.4	7.8
37	8.6	7.9
38	8.8	7.9
39	9.0	8.2
40	9.2	8.5
41	9.4	8.5
42	9.6	8.9
43	9.9	8.9
44	10.1	9.2
45	10.4	9.2
46	10.6	9.2

Table 6

Relationship Between Grade Equivalent Scores

Raw Score on Level 13	Level 13	Level 11
0	1.8	1.0
1	1.9	1.3
2	2.0	1.9
3	2.1	2.1
4	2.2	2.8
5	2.3	3.3
6	2.6	3.8
7	2.8	4.2
8	3.1	4.4
9	3.4	4.8
10	3.7	4.9
11	4.1	5.2
12	4.4	5.4
13	4.8	5.5
14	5.1	5.6
15	5.5	5.7
16	5.8	5.8
17	6.0	6.0
18	6.3	6.1
19	6.5	6.2
20	6.7	6.3
21	6.9	6.5
22	7.1	6.5
23	7.3	6.6
24	7.4	6.8
25	7.6	7.0
26	7.8	7.0
27	7.9	7.2
28	8.1	7.2
29	8.2	7.4
30	8.4	7.4
31	8.5	7.6
32	8.6	7.6
33	8.8	7.9
34	9.0	7.9
35	9.1	8.2
36	9.3	8.2
37	9.5	8.2
38	9.7	8.5
39	9.9	8.5
40	10.1	8.5
41	10.3	8.5
42	10.6	8.9
43	10.8	9.2
44	11.0	9.2
45	11.2	9.2
46	11.4	9.2
47	11.6	9.2
48	11.9	9.2

Table 7

Relationship Between Grade Equivalent Scores

Raw Score on Level 14	Level 14	Level 11
0	2.3	1.0
1	2.4	1.7
2	2.5	2.6
3	2.6	3.3
4	2.7	4.0
5	2.8	4.4
6	3.0	4.8
7	3.3	5.0
8	3.7	5.3
9	4.1	5.3
10	4.5	5.6
11	4.9	5.7
12	5.2	5.8
13	5.6	6.0
14	6.0	6.1
15	6.3	6.2
16	6.6	6.3
17	6.9	6.5
18	7.1	6.6
19	7.4	6.6
20	7.6	6.8
21	7.8	7.0
22	8.0	7.0
23	8.1	7.2
24	8.3	7.2
25	8.5	7.4
26	8.6	7.4
27	8.7	7.6
28	8.9	7.6
29	9.0	7.9
30	9.2	7.9
31	9.3	7.9
32	9.5	8.2
33	9.7	8.2
34	9.9	8.2
35	10.1	8.2
36	10.3	8.5
37	10.7	8.5
38	11.0	8.5
40	11.2	8.9
41	11.5	8.9
42	11.8	8.9
43	12.0	9.2
44	12.2	9.2
45	12.4	9.2
46	12.6	9.2
47	12.8	9.2
48	12.9	9.2

narrows, one still finds fairly high discrepancies. More dramatic discrepancies can be shown by comparing the upper end levels 12, 13, and 14 with level 11. For example, a comparison of level 13 with level 11 shows that a student capable of getting 89.5% of the questions correct on level 13 would encounter a ceiling effect on level 11. The grade equivalent score the student would achieve on level 11 would be 9.2 as compared to 10.8 on level 13.

Examination of the lower extreme on levels 11 and 13 show that students whose abilities were at chance on level 13 (or 12 of 48 items correct), would receive a grade equivalent score of 4.4 while on level 11 they would receive a score of 5.4.

Therefore, although in general agreement the results indicate that there are some fairly serious local discrepancies that may lead to different interpretations of what actually may be the same result. It also points out the care which with the assignment of tests must be made. The major discrepancies occur at the extremes of the raw score scale. Further, as expected, the severity of the discrepancies increase the levels with distance between the levels. Therefore, if off-level testing is to be used and the grade equivalent scores on different levels are to be considered as a single scale, care must be taken in that assignment of test levels and be within the relevant range of the student's ability. If the levels are assigned within this range, the grade equivalent scores obtained on any of the levels would appear to be sufficiently close for most practical comparisons.

References

Wright, B.D. Solving measurement problems with the Rasch model. Journal of Educational Measurement, 1977, 14, 97-116.

Wright, B.D. and Mead, R.J. BICAL: Calibrating Items and Scales with the Rasch Model. Research Memorandum No. 23A, Statistical Laboratory, Department of Education University of Chicago, January, 1978.